Sinai, M.J., McCarley, J.S., Krebs, W.K. (1999). Scene recognition with infrared, low-light, and sensor-fused imagery. Proceedings of the IRIS Specialty Groups on Passive Sensors. Monterey, CA.

Approved for public release; distribution is unlimited.

Scene Recognition with Infrared, Low-Light, and Sensor-Fused Imagery

February 1999

Sinai, Michael J., McCarley, Jason S., Krebs, William K.

Naval Postgraduate School
Department of Operations Research
1411Cunningham Rd
Monterey, CA 93943

ABSTRACT

The goal of this study was to evaluate the information conveyed by single- and dual-band sensor imagery by assessing performance on a scene recognition task. An experiment tested immediate recognition for pictures following an initial brief viewing. Images were taken with an uncooled LIMIRS long-wave infrared sensor and a Fairchild image intensified low-light CCD, and were compared against fused, false-color images created by remapping both spectral bands into a two-dimensional color space (Scribner et al., 1996). In each trial, an image was presented for 100 msec and followed immediately by a 300-msec checkerboard mask. A second image, of the same or of a different sensor format, was then displayed and remained visible until a response was made. The observer's task was to indicate whether the first and second image depicted the same scene, regardless of which sensor format the scenes were displayed in. All possible permutations of sensor formats were used. It was hypothesized that color-fusion, which combined and potentially enhanced information contained in the single-band images, would allow more accurate scene recognition. Performance was best when the first and second images were presented in the same format. When format changed between the presentation of the two images, performance deteriorated, but more so when the second image was of a single band format. Format of the first image itself had little effect, indicating that the primary benefits of sensor fusion were in matching the content of the second image to a stored representation of the first, and not in processing the briefly viewed first image. These results suggest that fusion can allow information from multiple single-band sensors to be effectively combined and presented within a single image, within which component information remains perceptually accessible.

1.0 INTRODUCTION

In evaluating sensor-fused nighttime imagery against more traditional single-band night vision imagery, studies have generally employed tasks meant to closely mimic those an operator might perform in a real-world environment (e.g., target detection, target identification). However, a comprehensive understanding of sensor fusion's potential benefits should include more detailed knowledge of the perceptual representation afforded by various sensor formats. Ideally, fused imagery will present information from multiple single-band sources undegraded within a unitary display, and will augment this representation with emergent chromatic or spatial information derived from the contrast between component images. It is not obvious, though, that sensor fusion will invariably be thus beneficial. While single-band information may be contained within a fused, dual-band image, it may not be perceptible. Rather, details of one component image may obscure those of the other, leaving single-band information within the fused image degraded or invisible to a human observer (Toet & Walraven, 1996; Essock, et al, 1999b). Emergent information, similarly, may be insalient even if present. The artificial color mappings produced by some fusion algorithms, finally, will not generally produce imagery whose chromatic characteristics correspond in any intuitive or obvious way to those of a scene viewed under natural photopic illumination. To the degree that human perception relies on stored knowledge of objects' chromatic characteristics, therefore, false color may be disruptive of perceptual performance. For all these reasons, fused imagery might fail to enhance human performance, and could even elicit worse performance than would single-band imagery alone. Past psychophysical research has in fact been equivocal in determining what utility, if any, sensor fusion has for human performance. While some studies have found a significant advantage for fused imagery over single sensor imagery (e.g., Essock, Sinai, McCarley, Krebs, & DeFord, 1999a; Toet, Ijspeert, Waxman, & Aguilar, 1997; Waxman, et al., 1996), others have not (Steele & Perconti, 1997; Krebs, et al., 1998; Essock, et al., 1999b).

These discrepant results can be attributed to the differences in fusion algorithms tested in the various studies, and to differences in the psychophysical tasks employed. Illustrative experiments by Essock, et al. (1999b), tested human performance on different psychophysical tasks using several fusion algorithms, including the one used in the current study. Results indicated that benefits of sensor fusion were mediated both by the nature of the perceptual task and by fusion algorithm. Fusion through some algorithms improved image segmentation, and hence facilitated recognition of various regions that were within a scene. Fusion through another algorithm, however, was detrimental to performance on the same task. Conversely, none of the fusion algorithms tested seemed to effectively retain details within regions of homogeneous texture. When observers were asked to discriminate regions which differed only in the orientation of their internal texture elements, performance was best with simple single-band IR imagery. These results underscore the need for a wide range of tasks to be evaluated in the testing of the different image formats. Absent an index of the perceptually salient single-band and emergent information within a fused image, however, it may be difficult to reconcile results obtained with various fusion algorithms and psychophysical tasks. Because sensor fusion can aid perceptual performance either by presenting multiple sources of single-band information concurrently, or by deriving information not available within component images singly, a demonstration that fusion influences image quality may by itself provide little information as to how, specifically, fusion was beneficial. Similarly, evidence that fusion is disruptive of perceptual performance may not by itself indicate specifically what information was degraded or sacrificed by a fusion algorithm. Tests of operator performance in applied tasks, therefore, can provide ambiguous implications for development of fusion algorithms. A measure of the perceptible single-band and emergent information conveyed by the sensor-fused imagery might eliminate these ambiguities, lending greater meaning to a finding that sensor fusion affects operator performance.

The current study used a test of immediate scene recognition to assess the information shared by and unique to various renderings of a common distal scene. The psychophysical task employed asked observers to each trial view a briefly presented image of a nighttime scene, rendered in one of several single-band or sensor-fused dual-band formats, and then to determine whether a test image presented in the same or a different format depicted the same scene. Given the demands of this task, performance should be determined by the degree to which the sensor formats in which the scenes to be compared are rendered convey similar sorts of distal information. If two sensor formats tend to convey similar information, that is, observers should be able to determine easily whether images rendered in those formats depict the same scene. If two sensor formats tend to convey information about different aspects of the depicted stimulus, conversely, observers might be expected to perform poorly when asked to determine whether images rendered in those formats depict the same or different distal objects. Under these assumptions, it was hypothesized that fused imagery, which is meant to convey information derived from multiple single-band sources, should allow for easier matching against imagery of a different format than would single-band information. The usefulness of the emergent chromatic information provided by false color was investigated by comparing performance with the fused monochrome images with their false color counterparts.

2.0 METHODS

Observers- Eighteen students from the Naval Postgraduate School were recruited for this experiment. All had normal or corrected to normal acuity by self report, and all granted informed consent prior to participation. All observers were active duty military.

Apparatus- Stimuli were displayed by a VisionWorks computer graphics system (Vision Research Graphics, Inc., Durham New Hampshire; Swift, Panish and Hippensteele, 1997) on a Nanao Flexscan F2.21 monitor. The monitor had a resolution of 800 x 600 pixels, a frame rate of 98.9 Hz, and a maximum luminance of 100 cd/m^2 with luminance linearized by means of a look up table. Observers viewed the screen from 1.5 meters.

Stimuli- Stimuli were images collected at Fort AP Hill, VA using an uncooled LIMIRS long-wave infrared sensor and a Fairchild image intensified low-light CCD. The images were of various nighttime scenes from around the installation, including wooded areas, fields, roads, and buildings. Images were later spatially registered and 'fused' by combining both spectral bands into a two-dimensional color space through an algorithm developed at the Naval Research Laboratory (Scribner, et al., 1996). The fusion algorithm used assigns each pixel a color vector determined by the detected power in the registered LL and IR imagery, with pixels that differ in their values of combined IR and LL power being presented in different intensities, and pixels that differ in their ordinal emissivity and reflectivity being presented in different hues. From a principle components analysis of the set of pixels in IR/LL space, the first principal component direction is taken to correspond approximately to an illuminant/radiant intensity vector. Intensity is assigned to the correlated component (major axis) for each pixel, and color is assigned to the uncorrelated feature (minor axis). Pixel color is assigned by opposing LL intensity against IR intensity and assigning one hue (cyan) to pixels whose intensity is greater in the LL than the IR image, and another hue (red) to pixels whose intensity is greater in the IR than in the LL image. Thus, the resulting image is displayed in two hues of various saturations. This coding system produces false-color imagery in which hue directly indicates the ordinal relative intensities of emitted and reflected energy at each pixel, lending a potential advantage for some perceptual tasks (Scribner et al., 1996).

A total of six image formats were tested: single-band IR and LL formats, two color-fused formats, and two achromatic fused formats. One color-fused version of each scene was derived using IR imagery of white-hot polarity, and the other using IR imagery of black-hot polarity. Achromatic versions of these fused images were spatially identical to their chromatic counterparts, but were rendered in grayscale. Single-band IR was of white-hot polarity. Twenty scenes were each rendered in these 6 different formats. Posttimulus pattern masks were checkerboard patterns consisting of 5' squares randomly assigned values from a look up table comparable to that of the masked image. If the initial image was achromatic, then the squares in its checkerboard mask were randomly assigned values from the grayscale look up table. If the initial image was chromatic, then the squares in its checkerboard mask were randomly assigned values from the look up table corresponding to one of the chromatic images. This insured that patterns mask contained colors similar to those in the images with which they were used. All images and masks had dimensions of 625 x 400 pixels. The surrounding screen was kept at a constant 50 cd/m^2 throughout experiment.

Procedure- Subjects were instructed that they would view two images each trial, and that their task was to judge whether or not the second image was of the same scene as the first, disregarding image format. Each trial began with a 250 ms presentation of a fixation cross. The first image was then presented at fixation for 100 ms, and was followed immediately by a 300 msec pattern mask. After this, the second image was presented, and remained visible until the observer responded. Responses were indicated with a keypress ("1" to indicate the two scenes were the same and "2" to indicate the two scenes were different). Auditory feedback was given following incorrect responses. After completing ten practice trials for which data was discarded, each observer provided data for 5 blocks of 72 trials. With 6 format types, 36 pair-wise permutations of image format were possible. Each of these pairings appeared twice within a block, once with the two scenes identical and once with them different. The first scene presented each trial was chosen randomly from the full set of twenty possible scenes. The second scene, when different from the first, was randomly chosen from the remaining pool of 19 scenes. The order of trials within a block was randomly determined. The dependent measure of interest was accuracy.

3.0 RESULTS

The results for all combinations of sensor format are plotted in Figure 1. These data are plotted where the format of the first image is shown along the abscissa and the format of the second (test) image is shown in the legend. The ordinate indicates mean error rates for all subjects. Thus, for example, the second bar in this graph shows the mean error rate when a scene was displayed briefly in black hot color fused format and then the test image shown second was displayed in the achromatic fused black hot gray format. A repeated measures ANOVA did not find a significant main effect of format for the image displayed first, $F_{(5,85)} < 1$. This can be seen graphically in Figure 2 where error rates are plotted by format of the initial image, collapsed across format of the second image, and where it is clear that format of the initial image had little effect on performance. This suggests that images of sensor-fused format were neither encoded nor stored more efficiently than images of single-band format, and, furthermore, that color-fused images were encoded and stored no more efficiently than achromatic images. The main effect of format for the image shown second, however, was reliable, $F_{(5,85)} = 7.58$, $p < .001$. This effect is illustrated in Figure 3, where the error rates are collapsed across format of first image displayed and are plotted by format of the second image. Clearly, there were more errors when the second image displayed was of either IR or LL format, as compared to when the second image was of a dual-band format. A post-hoc test comparing mean performance for the four conditions with a second image of dual-band format to mean performance for the two conditions with a second image of single-band format confirmed that this effect was reliable, $F_{(1, 35)} = 49.51$, $p < .001$. The omnibus interaction of format of the first image by format of the second image was also significant, $F_{(25,425)} = 2.01$, $p = .003$.

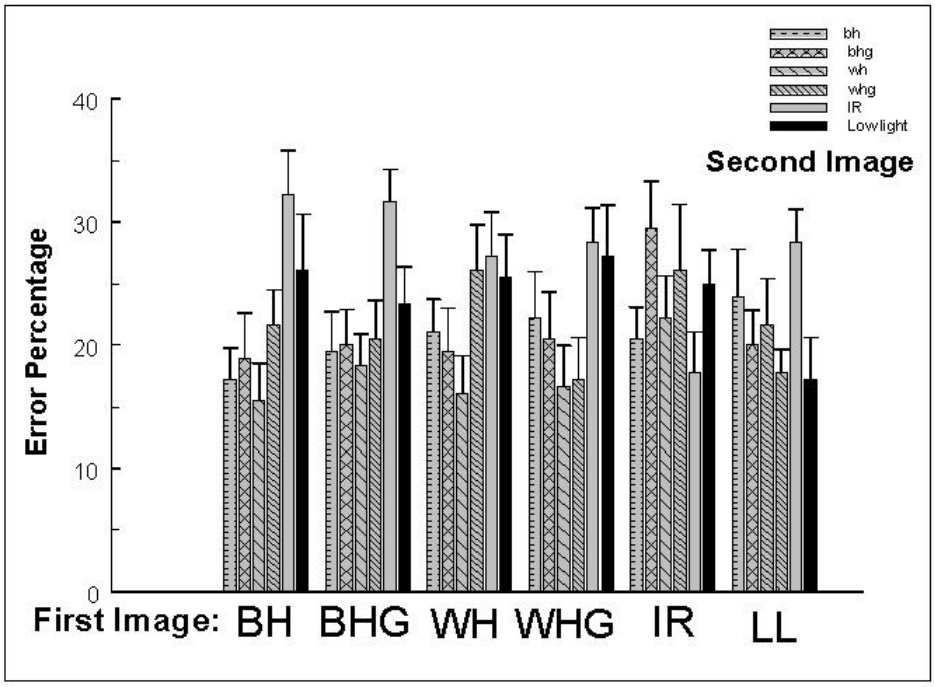

Figure 1. Error percentage plotted by sensor format type for both orders of presentation. The legend shows the key for which sensor format type was shown second. The sensor format type of the first image is plotted along the ordinate. BH=color fused black hot, BHG=achromatic fused black hot, WH=color fused white hot, WHG=achromatic fused white hot, IR=infrared, LL=low-light.

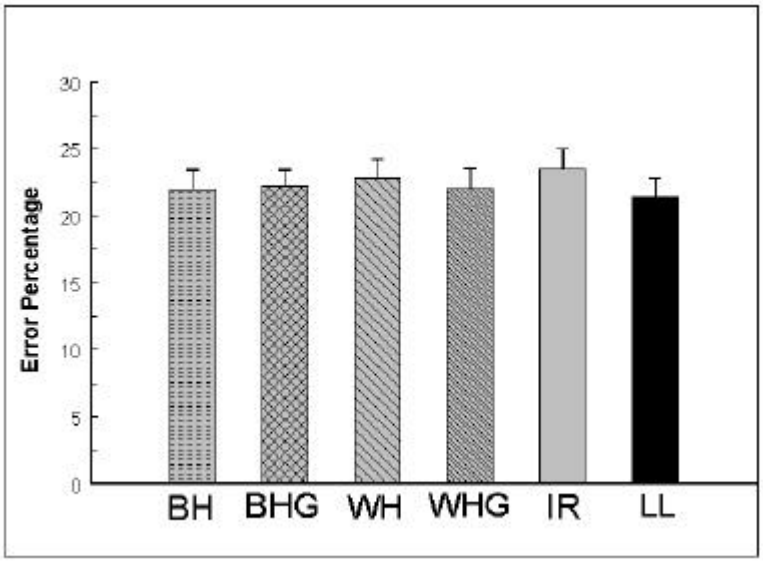

Figure 2. Error percentage plotted for the sensor format of the first image, collapsed over format of the second image. Error bars show +1 S.E.M .

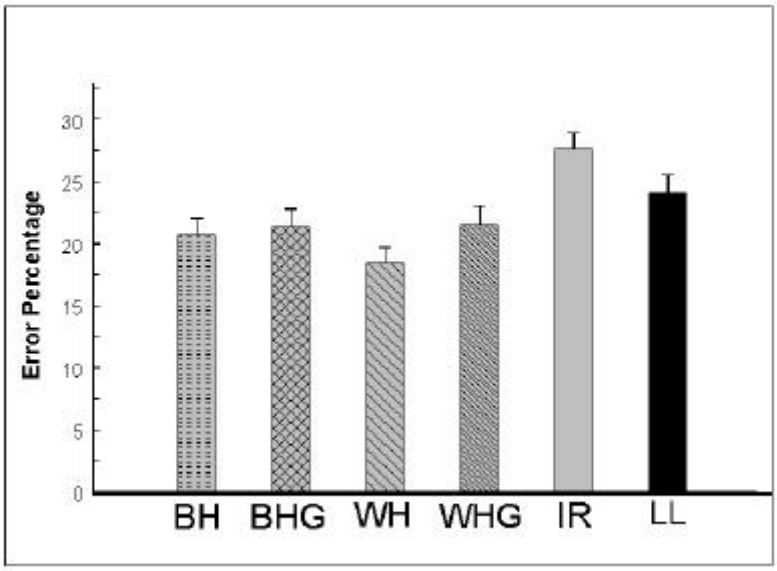

Figure 3. Error percentage plotted for the sensor format of the second image, collapsed over format of the first image. Error bars show +1 S.E.M .

This reliable interaction presumably arose largely from the tendency for error rates to increase when the second image presented on a given trial was of any format different from the first. An additional analysis was conducted, however, to preclude another potential explanation. It is possible that the reliable main effect of dual-band imagery in the omnibus ANOVA arose simply because dual-band images could be matched easily against other dual–band images, and that the significant interaction arose because fused imagery was no more easily matched against single-band images than were single-band images of different formats matched against one another. To test this possibility, data collected from conditions in which the format of the first image was the same as that of the second image was discarded, and individual image formats were collapsed into three groups: dual-band chromatic, dual-band achromatic, and single-band. Data are illustrated in Figure 4, where it is clear that dual-band imagery, presented for match against a stored representation, allowed consistently better performance than did single-band imagery. Data were submitted to a two-way ANOVA with format of first image (dual-band chromatic, dual-band achromatic, or single-band) and format of second image (dual-band chromatic, dual-band achromatic, or single-band) as factors. Results again indicated a reliable main effect of the second image's format, $F_{(2, 34)} = 17.09$, but failed to indicate reliable interaction, $F_{(4, 68)} < 1$. Thus, data suggest that a dual-band second image was more easily matched against a stored image of any format than was a single-band second image, $F_{(2, 34)} < 1$. Results again failed to indicate a reliable main effect of the first image's format. Although data suggest that performance was better when the second image presented on a trial was dual-band chromatic rather than dual-band achromatic, an additional two-way ANOVA that compared these two conditions using format of first image (dual-band chromatic or dual-band achromatic) and format of second image (dual-band chromatic or dual-band achromatic) as factors failed to produce any reliable effects, all p's > .12.

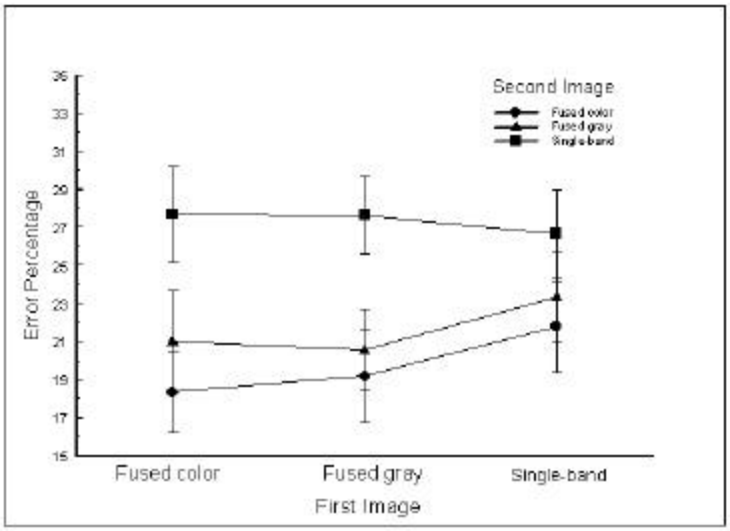

Figure 4. Mean error percentages for chromatic dual-band fused imagery, achromatic dual-band fused imagery, and single-band imagery.

4.0 DISCUSSION

The experimental task employed asked observers to view and remember a briefly presented image, then to determine whether or not a second image, presented 300 msec later, depicted the same distal scene. Recognition was best when the first and second images were presented in the same format. The adverse effect of changing sensor format between the two scenes, however, was greater when the second image was displayed in a single-band rather than sensor-fused format. The format of the first image itself had little effect at all, suggesting that the primary benefits of sensor fusion in this task were in matching the content of the second image to a stored representation of the first, and not in processing the briefly viewed first image. These results suggest that fused images retained perceptually accessible information from the original images. That is, observers in this study responded as if they could still access the original IR and LL images from the fused imagery, even after the original images had been modified and rendered in color. The poor performance that occurred when the second image was in a single-band format different from that of the first image suggests that information could not be transferred between the single sensor imagery. Thus fusion appeared to allow information from multiple single-band sensors to be effectively combined and presented within a single image, wherein it remained perceptually accessible.

These results appear at first examination to conflict with those of an earlier study employing the same fusion algorithm. As noted in the introduction, Essock, et al (1999b), found that fusion of IR and LL images through the current algorithm appeared to compromise texture information, impairing performance on a psychophysical task that required observers to discriminate regions of texture elements that varied only in their orientation. The apparently discrepant results between that study and the current experiment can most likely be attributed to differences in the information required to perform the different psychophysical tasks employed. In the current experiment, observers were asked to view and compare images of full scenes. In the study reported by Essock, et al, observers could perform their task only on the basis of low-level texture information, the spatial microdetail within an image (Essock, 1992). The apparent utility of fusion in the present experiment, and its apparent detriment to performance in the

previous study, may indicate simply that image fusion effaced textural details but retained perceptual contrast between large-scale regions of different textures. Essock, et al (1999b), comparing the effects of fused imagery on different psychophysical tasks, reached a similar conclusion. The current experimental task, notably, provides a direct method of testing these notions, and more generally, of measuring how well information of various types and at various spatial scales is conveyed by fused imagery.

An unexpected finding among the current results was that while performance was moderated by format of the second image presented each trial, there was little effect of the first image's format. This is perhaps especially surprising given that the first image presented each trial was displayed only briefly (100 msec) and then masked. Under these circumstances, image quality might have been expected to effect a large influence on perceptual performance, and in particular, chromatic information might have been expected to facilitate image segmentation and encoding (Gegenfurtner, 1997). The failure of image format to influence results under these circumstances, however, might indicate only that the duration for which images were exposed was not sufficiently brief to hinder perceptual encoding. Past research, in fact, has shown that observers can recognize objects and encode the meaning of a scene in a glimpse of far less than 100 msec duration (Biederman, Rabinowitz, Glass, & Stacy, 1974; Potter, 1976). Briefer exposure durations in the current study might have revealed more obvious effects of image format on the ease with which images were segmented and encoded, and perhaps clearer benefits of chromatic imagery. Further research will be necessary to test these predictions.

REFERENCES

Biederman, I. Rabinowitz, J. C., Glass, A. L., and Stacy, E. W. (1974). On the information extracted from a glance at a scene. Journal of Experimental Psychology, 103, 597-600.

Essock, E. A. (1992). An essay on texture: The extraction of stimulus structure from the visual image. In B. Burns (Ed.), Percepts, Concepts, and Categories: The Representation and Processing of Visual Information. Amsterdam, Holland: North-Holland.

Essock, E. A., Sinai, M. J., McCarley, J. S. and Krebs, W. K., and DeFord, J. K. (1999a) Perceptual ability with real-world nighttime scenes: Image-intensified, infrared and fused-color imagery. Human Factors, in press.

Essock, E. A., Krebs, W. K., Sinai, M. J., DeFord, J. K., Srinivasan, N., and McCarley, J.S.(1999b) Human Perceptual Performance With Nighttime Imagery: Region Recognition and Texture-Based Segmentation. Manuscript in preparation.

Gegenfurtner, K. R. (1997) Color in the recognition of natural scenes. Investigative Ophthalmology and Visual Sciences, Suppl., 38, S900.

Krebs, W. K., Scribner, D. A., Miller, G. M., Ogawa, J. S., Schuler, J. (1998). Beyond third generation: a sensor fusion targeting FLIR pod for the F/A-18. Proceedings of the SPIE-Sensor Fusion: Architectures, Algorithms, and Applications II, 3376, 129-140.

Potter, M. C. (1976). Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory, 2, 509-522.

Scribner, D. A., Satyshur, M. P., Schuler, J. and Kruer, M. P. (1996) Infrared color vision. IRIS Specialty Group on Targets, Backgrounds and Discrimination, January, Monterey, CA.

Steele, P. M. and Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. SPIE 11th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando FL.

Swift, D. J., Panish, S. and Hippensteel, B. (1997) The use of VisionWorks in visual psychophysics research. Spatial Vision, 10, 471-477.

Toet, A., Ijspeert, J. K., Waxman, A. M., and Aguilar, M. (1997) Fusion of visible and thermal imagery improves situational awareness. Proceedings of the SPIE Conference on Enhanced and Synthetic Vision 1997, SPIE-3088, 177-188.

Toet, A. & Walraven, J. (1996) New false color mapping for image fusion. Optical Engineering, 35, 650-658.

Waxman, A. M., Gove, A. N., Seibert, M. C., Fay, D. A., Carrick, J. E., Racamato, J. P., Savoye, E. D., Burke, B. E., Reich, R. K., McGonagle, W. H. and Craig, D. M. (1996). Progress on color night vision: Visible/IR fusion, perception and search, and low-light CCD imaging. Proceedings of the SPIE Conference on Enhanced and Synthetic Vision, SPIE-2736, 96-107.